National Digital Strategy Advisory Board
Report from June 28, 2004 Meeting

The June 28, 2004, meeting of the National Digital Strategy Advisory Board offered the promise of a new world of digital preservation while acknowledging the great challenges that lie ahead.

"Information technology itself is moving ahead with unpredictable leaps. Inevitably, it offers us new topics as well as new challenges," said Librarian of Congress James H. Billington. "The degree to which our shared knowledge and cultural heritage are in digital form is, of course, greater than ever.

"But what is most interesting," he continued, "and really exciting about our meeting today, is that it will feature reports of excellent progress in numerous digital preservation efforts in a variety of institutions."

The Librarian was encouraged by such projects as the soon-to-be-awarded cooperative agreements with eight institutions to collect and preserve valuable digital content. He praised "the power of collaboration" with such organizations as the International Internet Preservation Consortium and the Digital Preservation Coalition in the United Kingdom.

Dr. Billington then announced that the Library had just entered into an agreement with the National Science Foundation to fund advanced research in the preservation of digital materials.

An overview of the day's agenda was then delivered by Laura E. Campbell, associate librarian for Strategic Initiatives, who is leading the National Digital Information Infrastructure and Preservation Program (NDIIPP) for the Library of Congress.

She noted that NDSAB members would hear about the important digital preservation work of the National Library of Medicine, the National Archives and Records Administration, the National Agricultural Library and the Institute for Museum and Library Services. The afternoon would be devoted to an update on the progress of NDIIPP and a discussion of salient copyright issues.

Campbell thanked all the board members for their help in advancing the field of digital preservation. "It's made a huge difference having so many people who have real dedication to this work help guide us and share in some of the decision-making," she said.

The meeting was an occasion to welcome two new members to the board: Bruce James, the Public Printer of the United States, and the Institute for Museum and Library Services, represented by Joyce Ray, associate deputy director for library services. Betsy Humphreys, associate director of library operations and assistant director for health services research information at the National Library of Medicine (NLM), offered a "Digital Preservation Update (PDF)" for her agency.

She noted that the preservation of digital materials must necessarily be a part of NLM's mission, just as it is the mission of the Library of Congress and other agencies to preserve materials in all formats. "We are focused on the notion of permanent access to electronic information and the idea that we should use NLM's own electronic output and its services as testbeds for pushing forward in digital preservation. And we should

definitely work with the other national libraries and other relevant governmental and nongovernmental organizations to develop a strategy that will be scalable and eventually national."

One of the highlights of her discussion was a demonstration of NLM's Web archive. The archive offers "documents that we want to keep because they're part of the NLM archives, but, in fact, are no longer current." The archive allows users to view the current iteration of a particular document as well as pervious versions.

Taking a cue from NDIIPP's cooperative approach, Humphreys said that "our thought is to see if we can … export this approach to other government agencies."

Before introducing the next speaker, Campbell expressed her satisfaction that the meeting was turning out to be "an opportunity to see some of the amazing good work and strengths that we have in this group as a whole, as we lean on one another to leverage this great work."

Kenneth Thibodeau, director of the Electronic Records Archives Program at the National Archives and Records Administration (NARA), mentioned that in his presentation (PDF) he would "talk a little bit about the peculiarities you get when you're dealing with records, as opposed to publications and creative works."

Thibodeau told the audience that "we're only at the beginning of the digital era. We have to anticipate that 10 to 20 years from now there are going to be all kinds of digital information that has barely been dreamt of. … If you're going to build a system that's going to last for a while, you'd better build it in a way that can continue to take advantage of improvements in the technology." His point was especially relevant for NDIIPP, as one of the major areas of investment for the program will be in continuing research, so that digital materials remain compatible with advances in technology.
For an agency such as NARA, which is responsible for preserving and making available the records of myriad government agencies that often use different storage methods, the task is daunting. "So you want your system to be able to evolve and you also want it basically to be transparent. … [Users need to be confident that] they're getting authentic records, not that they're dealing with a NARA-specific system. So we make similar assumptions, that our system must be able to interact with very different systems."

The Archives is testing a distributed archives that exists at its facility in College Park, Md.; at the University of Maryland's Institute for Advanced Computer Systems; and at the San Diego Super Computer Center.

"What we've done," said Thibodeau, is to take chunks of NARA's existing holdings of electronic records … and we've distributed them across these three sites, each of which uses different operating systems and different hardware. But [for] a user logging onto the archive, that's all transparent."

Judith Russell, Superintendent of Documents for the Government Printing Office, focused on GPO's Federal Depository Library Program, which is responsible for the distribution and preservation of thousands of government publications. One of the primary methods of GPO distribution is its Access System, which makes content available electronically and celebrated its 10th anniversary in 2004.

"Our depository program is becoming digital … and it's completely transforming the way we do business," Russell said.

"One of the things that Laura [Campbell] had posed to us in making presentations was to discuss collaboration. This is probably the ultimate project for collaboration. The

collections are in the depositories. Many of the federal agencies don't even have as good collections of their own publications as the depository libraries have," she said.

Russell noted hopefully, however, that "there's an enormous interest and willingness in the depository community to digitize these materials for public access. They're extremely committed to the idea that these are public domain resources and that in digitizing them, they should be maintained as public domain resources." Russell said she believes that, working collaboratively, GPO can digitize and make publicly available its collection of publications "in the next three to five years."

The National Agricultural Library (NAL) was represented by Eleanor Frierson, the deputy director. She pointed out in her presentation (PDF) that "historically, we have digitized for access. We haven't necessarily digitized for preservation. The National Agricultural Library, which is in Beltsville, Md., focused on getting the information out to its customers nationally and internationally using technology."

She also noted that collaboration has been a part of NAL's mode of operation because "it's been recognized for a very, very long time that agriculture is a global enterprise and that similar crops grow in similar zones across the world." … So within the agriculture community, there was a great deal more collaboration than there was in other communities."

NAL is in the beginning stages of launching a national digital library for agriculture. "And we are absolutely looking to NDIIPP as a very important way to help us set standards and best practices for preservation. … We want to get it done and do it right the first time with a minimum of agony."

A primary example of collaboration in the agriculture community is the United States Agricultural Information Network (USAIN), a voluntary network of U.S. libraries, both public and private, as well as international partners. USAIN members are working to develop a preservation program for agricultural literature with an emphasis on pre-1950 U.S. publications. Frierson said she was "gratified" that "the number of states participating in this will soon be 27."

Another program, the USDA Publications Preservation Program, or DP3, is working on a plan to preserve born-digital publications of the Department of Agriculture. "Through the DP3 work, we identified a bunch of issues that are not a surprise to anyone here." Her next statement confirmed what was learned during the initial phases of NDIIPP: that technology is not the greatest barrier to digital preservation.

"It's the people stuff that's tough, not necessarily the technology stuff," Frierson said. "It's the people and the issues that change."

The Institute of Museum and Library Services (IMLS), an independent federal grant-making agency dedicated to helping libraries and museums serve their communities, admittedly "doesn't preserve anything other than its own internal records," said Joyce Ray, associate deputy director for library services. "But we, of course, are very concerned about what happens to our [grant] money and particularly the money that goes into digitization."

In her presentation (PDF), Ray reminded the audience that her agency is young, having been created by Congress in 1996 and making its first competitive grants to libraries in 1998. "We do have the distinction, I believe, of being the only federal-funding agency that has statutory authority for digitization, and I think that is because we are a

new agency, so it was actually written into our authorization. So we really had a mandate to fund digitization.

"There has been so much interest and demand for funds for digitization that we couldn't have avoided it even if we had wanted to, even in 1998 when we gave our first competitive grants. But many questions were raised about preservation that are just now being addressed in a more comprehensive way," Ray continued. "We have all learned a lot from the digital projects we have funded over the last several years."

IMLS's National Leadership Grant Program funds both digital conversion projects as well as creation of new content such as learning modules and research and demonstration projects "that get at more issues like research to develop tools and standards and to investigate approaches to preservation."

Ray mentioned that the University of Illinois at Urbana-Champaign (an NDIIPP partner) "has taken the lead for us by directing a project that is testing the use of the OAI metadata harvesting protocol using IMLS-funded digital collections as a testbed, and they are developing a collection-level registry as well as an item-level search engine." (OAI, the Open Access Initiative, develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content.)

In the area of digital preservation, IMLS funded two significant projects in 2002. The California Digital Library, in partnership with the University of California Berkeley Library, is creating a model preservation repository for multi-institutional digital materials following the Open Archival Information System (OAIS) reference model. The University of Florida's Florida Center for Library Automation is developing a model central digital archiving facility for the libraries of Florida's public college and university system, also based on OAIS. The two projects differ in that Florida is developing a "dark archive" and is looking especially at cost-recovery issues associated with archiving, while California is creating an open repository and evaluating user needs and interfaces.

IMLS is in the process of doing "some high-level thinking about what will be the needs for digital libraries, both within libraries and museums, and potentially in new types of organizations, in the future."

The afternoon session began with an NDIIPP progress report (PDF) by Campbell. She affirmed what many of the morning's speakers had emphasized: "The key area here is cooperation with others. … Again, as we keep hearing, there's a theme that the human organizational dynamic may be one of the harder" hurdles to overcome.

She offered a brief overview of the steps that led to Congress's approval of the so-called "master plan" for NDIIPP, formally called "Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program." "We now know through our early discussions that there are really two major components for an overall national preservation infrastructure: … the network of partners and the technical architecture."

From the beginning of planning for NDIIPP until the NDSAB meeting, Campbell was pleased to report that 12 terabytes of at-risk digital material had been identified for preservation. She was also happy to report that the Library had received 22 qualified proposals to its Program Announcement to Support Building a Network of Partners. (Eight winners were selected in September 2004.)

"We have a huge opportunity here, and that opportunity is to document this experience in such a way that we can communicate some of the huge challenges that we have before us, not the least of which is intellectual property."

Before that issue could be addressed by Register of Copyrights Marybeth Peters, Clay Shirky, an adjunct professor in New York University's graduate Interactive Telecommunications Program and a consultant to NDIIPP, spoke in his presentation (PDF) about the Archive Ingest and Handling Test (AIHT). AIHT is testing the intake (or "ingest") of an archive into diverse systems. The participants will also work to understand the difficulties in transferring large and complex digital archives from one institution to another. This is a critical piece of any larger digital preservation effort, as the number of individuals and organizations that produce digital material is far larger, and growing much faster, than the number of institutions committed to preserving such material. Thus, any practical preservation strategy requires mechanisms for continuous transfer of content from the wider world into the hands of preserving institutions.

Shirky warned those looking for a simple solution for digital preservation that "the number of places that are taking on library-like functions is increasing radically. … And the number of organizations doing things like specifying content management systems or dealing with refreshing of bits is too large for there ever to be one right system."

Turning an aphorism on its head, Shirky said that in the world of digital preservation, "great minds do not think alike. The more institutions we talk to, the more models we find for long-term preservation."

Thus, the goal of the NDIIPP architecture "is a lens for us. It's not a template for building a particular system but, rather, it lets us evaluate existing systems."

Shirky was followed by Larry Brandt, who manages the Digital Government Research Program at the National Science Foundation (NSF). Brandt spoke of a memorandum of understanding between the Library of Congress and NSF to establish the first research grants program to specifically address digital preservation. NSF will administer the program, which will fund cutting-edge research to support the long-term management of digital information. He emphasized in his presentation (PDF) that NSF would be looking for "high-risk" proposals; "that is, things that might not work, not incremental extensions to current thinking. And we're hoping that these will be innovative enough that they might catalyze advances in this interdisciplinary research area." Grants were to be announced during spring 2005.

Campbell introduced Peters as one who would discuss intellectual property issues in the digital world, but Peters began, saying, "I'm actually going to broaden it a little bit. "When NDIIPP started, the focus was on preserving materials, so it was a preservation program [exclusively]. … But in order to preserve something, you have to have it. So acquisition became part of the issue that we have to look at."

Acquiring digital materials presents a set of problems that do not exist in the analog world. For example, "when we talk about Web sites that [statistically] disappear in 44 days, asking people to send those to you [under copyright depository requirements] is really not a very good way of acquiring that material," Peters said, noting that the Library would be working with Congress to suggest changes to copyright laws and would form an advisory group to study the problems.

She closed her discussion (PDF) by telling the audience that “you have a fantastic opportunity to make a difference, to document what the issues are and to bring about changes that could benefit all of the people of the United States.”

With that, the presentations and the NDSAB meeting drew to a close. Campbell remarked “how grateful I am to each and every one of you and all the contributions that you’ve made over the last couple of years to helping us move along. I’m particularly pleased to hear about all the good work being done.”

Billington, echoed Campbell, saying, “how positive this group has been, how helpful you are and how increasingly important our work is going to be” as the National Digital Information Infrastructure and Preservation Program moves forward. “Thank you for what you have done, and we look forward to seeing you again soon and hearing from you in the meantime on this important new enterprise.”